

Network Work group
Internet-Draft
Intended status: Standards Track
Expires: September 6, 2015

N. Kumar
C. Pignataro
N. Akiya
Cisco Systems, Inc.
L. Zheng
M. Chen
Huawei Technologies
G. Mirsky
Ericsson
March 5, 2015

BIER Ping and Trace
draft-kumarzheng-bier-ping-00

Abstract

Bit Index Explicit Replication (BIER) is an architecture that provides optimal multicast forwarding through a "BIER domain" without requiring intermediate routers to maintain any multicast related per-flow state. BIER also does not require any explicit tree-building protocol for its operation. A multicast data packet enters a BIER domain at a "Bit-Forwarding Ingress Router" (BFIR), and leaves the BIER domain at one or more "Bit-Forwarding Egress Routers" (BFERs). The BFIR router adds a BIER header to the packet. The BIER header contains a bit-string in which each bit represents exactly one BFER to forward the packet to. The set of BFERs to which the multicast packet needs to be forwarded is expressed by setting the bits that correspond to those routers in the BIER header.

This document describes the mechanism and basic BIER OAM packet format that can be used to perform failure detection and isolation on BIER data plane.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 6, 2015.

Copyright Notice

Copyright (c) 2015 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
2. Conventions used in this document	3
2.1. Terminology	3
2.2. Requirements notation	3
3. BIER OAM	4
3.1. BIER OAM message format	4
3.2. Return Code	6
3.3. BIER OAM TLV	7
3.3.1. Original SI-BitString TLV	7
3.3.2. Target SI-BitString TLV	8
3.3.3. Incoming SI-BitString TLV	9
3.3.4. Downstream Mapping TLV	10
3.3.5. Responder BFER TLV	12
3.3.6. Responder BFR TLV	13
3.3.7. Upstream Interface TLV	14
4. Procedures	14
4.1. BIER OAM processing	14
4.2. Per BFER ECMP Discovery	15
4.3. Sending BIER Echo Request	15
4.4. Receiving BIER Echo Request	16
4.5. Sending Echo Reply	17
4.6. Receiving Echo Reply	17
5. IANA Considerations	17
5.1. Message Types, Reply Modes, Return Codes	17
5.2. TLVs	17
6. Security Considerations	18
7. Acknowledgement	18
8. Contributing Authors	18
9. References	18

9.1. Normative References	18
9.2. Informative References	19
Authors' Addresses	19

1. Introduction

[I-D.wijnands-bier-architecture] introduces and explains BIER architecture that provides optimal multicast forwarding through a "BIER domain" without requiring intermediate routers to maintain any multicast related per-flow state. BIER also does not require any explicit tree-building protocol for its operation. A multicast data packet enters a BIER domain at a "Bit-Forwarding Ingress Router" (BFIR), and leaves the BIER domain at one or more "Bit-Forwarding Egress Routers" (BFERs). The BFIR router adds a BIER header to the packet. The BIER header contains a bit-string in which each bit represents exactly one BFER to forward the packet to. The set of BFERs to which the multicast packet needs to be forwarded is expressed by setting the bits that correspond to those routers in the BIER header.

This document describes the mechanism and basic BIER OAM packet format that can be used to perform failure detection and isolation on BIER data plane without any dependency on other layers like IP layer.

2. Conventions used in this document

2.1. Terminology

BFER - Bit Forwarding Egress Router

BFIR - Bit Forwarding Ingress Router

BIER - Bit Index Explicit Replication

ECMP - Equal Cost Multi-Path

OAM - Operation, Administration and Maintenance

SI - Set Identifier

2.2. Requirements notation

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

3. BIER OAM

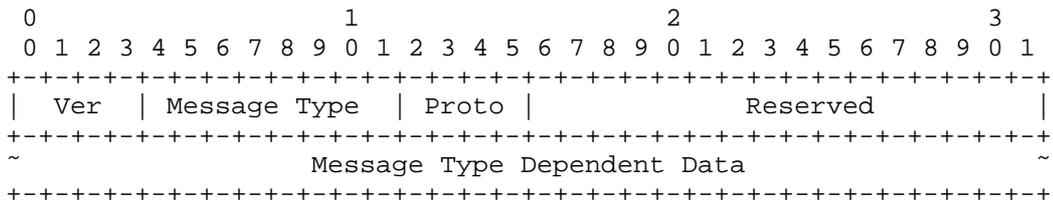
BIER OAM is defined in a way that it stays within BIER layer by following directly the BIER header without mandating the need for IP header. [I-D.wijnands-mpls-bier-encapsulation] defines a 4-bit field as "Proto" to identify the payload following BIER header. In order to differentiate the BIER data packet from BIER OAM packet, this document introduces a new value for the Proto field as:

Proto:

PROTO-TBD1: BIER OAM

3.1. BIER OAM message format

The BIER OAM packet header format that follows BIER header is as follows:

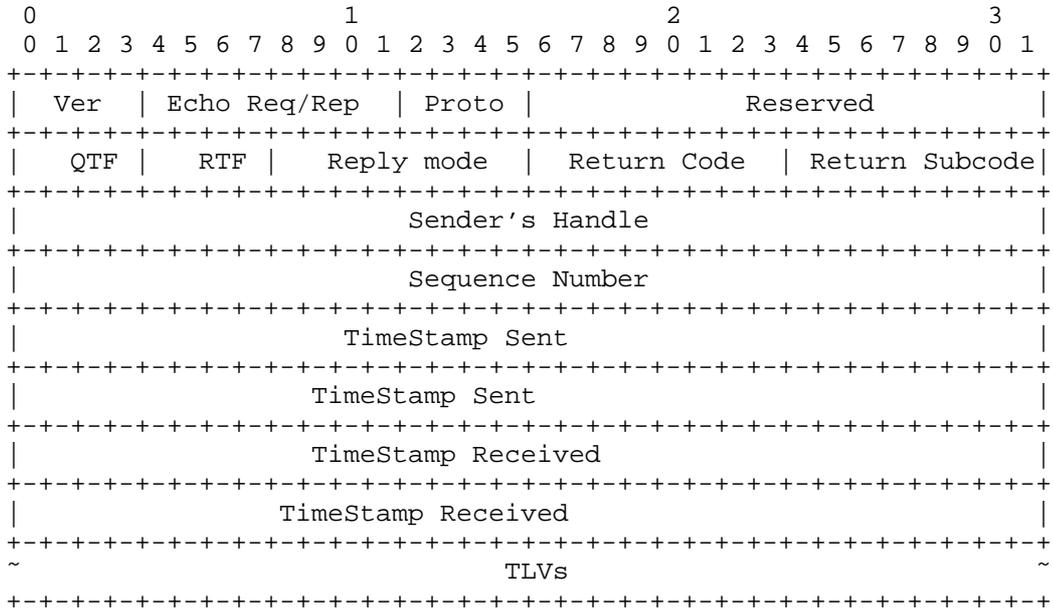


Type

The Message Type is one of the following:

Type	Value Field
TBD1	BIER Echo Request
TBD2	BIER Echo Reply

The Echo Request/Reply header format is as follows:



Proto

Set to 0 for Echo Request/Reply header.

QTF

Querier Timestamp Format. When set to 2, the Timestamp Sent field is (in seconds and microseconds, according to the Initiator's clock) in NTP format [RFC5905]. When set to 3, the timestamp format is in IEEE 1588-2008 (1588v2) Precision Time Protocol format.

RTF

Responder Timestamp Format. When set to 2, the Timestamp Received field is (in seconds and microseconds, according to the Initiator's clock) in NTP format [RFC5905]. When set to 3, the timestamp format is in IEEE 1588-2008 (1588v2) Precision Time Protocol format.

Reply mode

The Reply mode is set to one of the below:

Value	Meaning
1	Do not Reply
2	Reply via IPv4/IPv6 UDP packet.
3	Reply via BIER packet

Return Code

Set to zero if Type is TBD1. Set as defined in section 3.2 if Type is TBD2.

Return subcode

To Be updated.

Sender's Handle, Sequence number and Timestamp

The Sender's Handle is filled by the Initiator, and returned unchanged by responder BFR. This is used for matching the replies to the request.

The Sequence number is assigned by the Initiator and can be used to detect any missed replies.

The Timestamp Sent is the time when the echo request is sent. The TimeStamp Received in echo reply is the time (accordingly to responding BFR clock) that the corresponding echo request was received. The format depends on the QTF/RTF value.

TLVs

Carries the TLVs as defined in Section 3.3.

3.2. Return Code

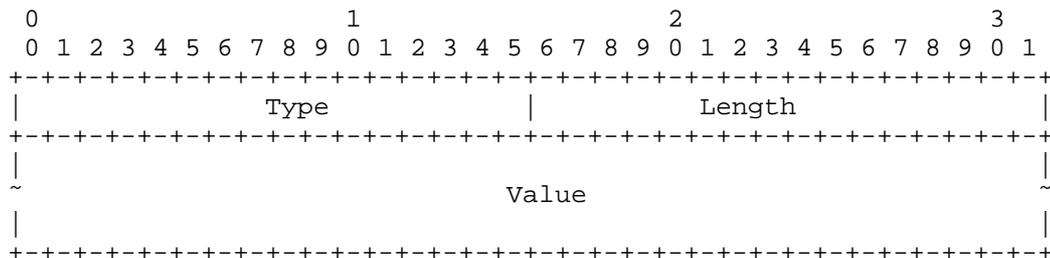
Responder uses Return Code field to reply with validity check or other error message to Initiator. It does not carry any meaning in Echo Request and MUST be set to zero.

The Return Code can be one of the following:

Value	Value Meaning
0	No return code (Set by Initiator in Echo Request)
1	Malformed echo request received
2	One or more of the TLVs was not understood
3	Replying BFR is the only BFER in header Bitstring
4	Set-Identifier Mismatch
5	Packet-Forward-Success
6	Invalid Multipath Info Request
7	No control plane entry for Multicast Overlay Data
8	No matching entry in forwarding table.
9	Replying BFR is one of the BFER in header Bitstring

3.3. BIER OAM TLV

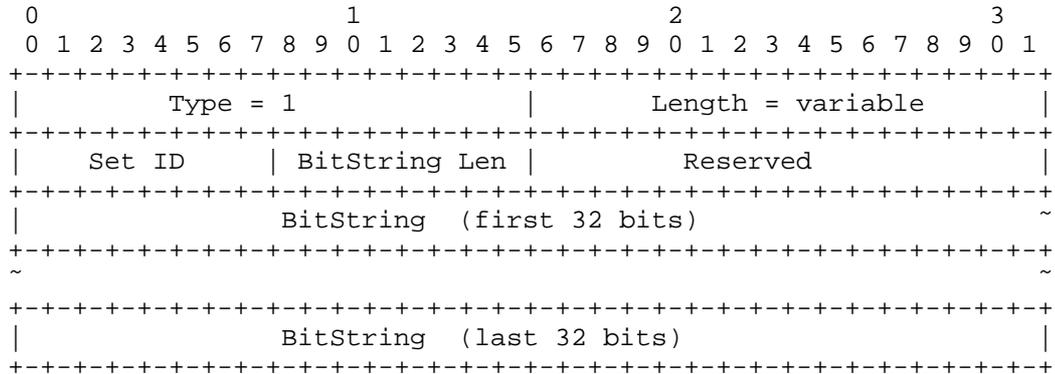
This section defines various TLVs that can be used in BIER OAM packet. The TLVs (Type-Length-Value tuples) have the following format:



TLV Types are defined below; Length is the length of the Value field in octets. The Value field depends on the TLV Type.

3.3.1. Original SI-BitString TLV

The Original SI-BitString TLV carries the set of BFER and carries the same BitString that Initiator includes in BIER header. This TLV has the following format:



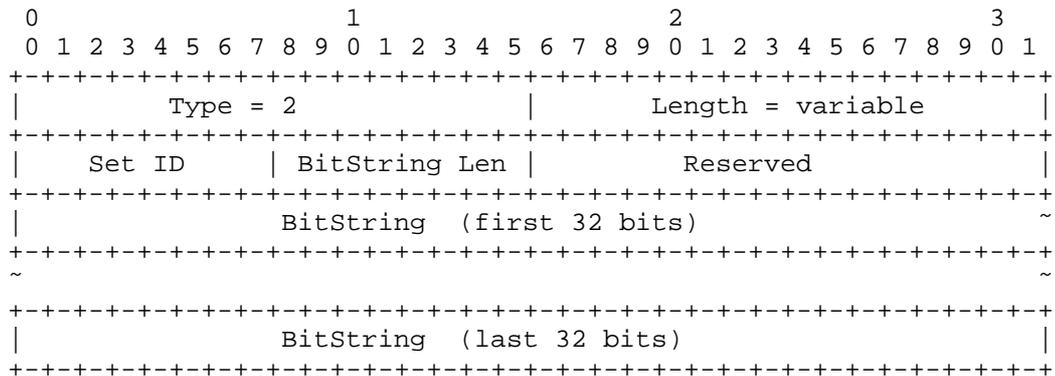
The Length field is set as defined in section 3 of [I-D.wijnands-mpls-bier-encapsulation].

Set ID field is set to the Set Identifier to which the BitString belongs to. This value is derived as defined in section 3 of [I-D.wijnands-bier-architecture]

The BitString field carries the set of BFR-IDs that Initiator will include in the BIER header. This TLV MUST be included by Initiator in Echo Request packet

3.3.2. Target SI-BitString TLV

The Target SI-BitString TLV carries the set of BFER from which the Initiator expects the reply from. This TLV has the following format:



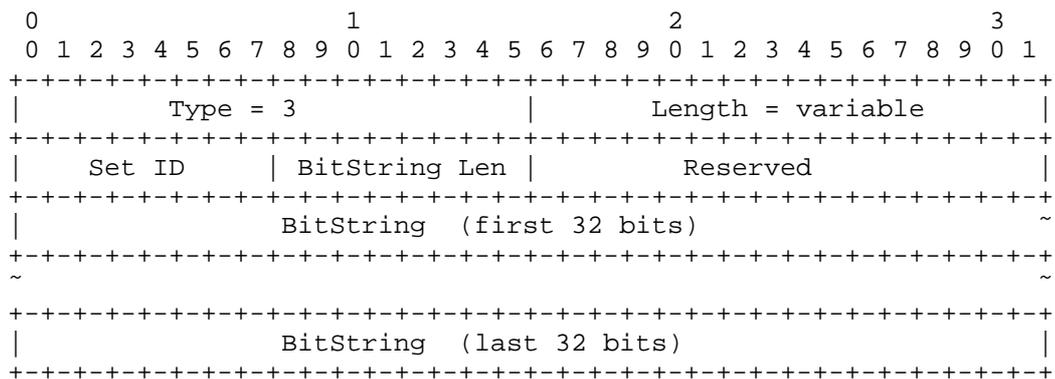
The Length field is set as defined in section 3 of [I-D.wijnands-mpls-bier-encapsulation].

Set ID field is set to the Set Identifier to which the BitString belongs to. This value is derived as defined in section 3 of [I-D.wijnands-bier-architecture]

The BitString field carries the set of BFR-IDs of BFER(s) that Initiator expects the response from. The BitString in this TLV may be different from the BitString in BIER header and allows to control the BFER responding to the Echo Request. This TLV MUST be included by Initiator in BIER OAM packet if the Downstream Mapping TLV (section 3.3.4) is included.

3.3.3. Incoming SI-BitString TLV

The Incoming SI-BitString TLV will be included by Responder BFR in Reply message and copies the BitString from BIER header of incoming Echo Request message. This TLV has the following format:



The Length field is set as defined in section 3 of [I-D.wijnands-mpls-bier-encapsulation].

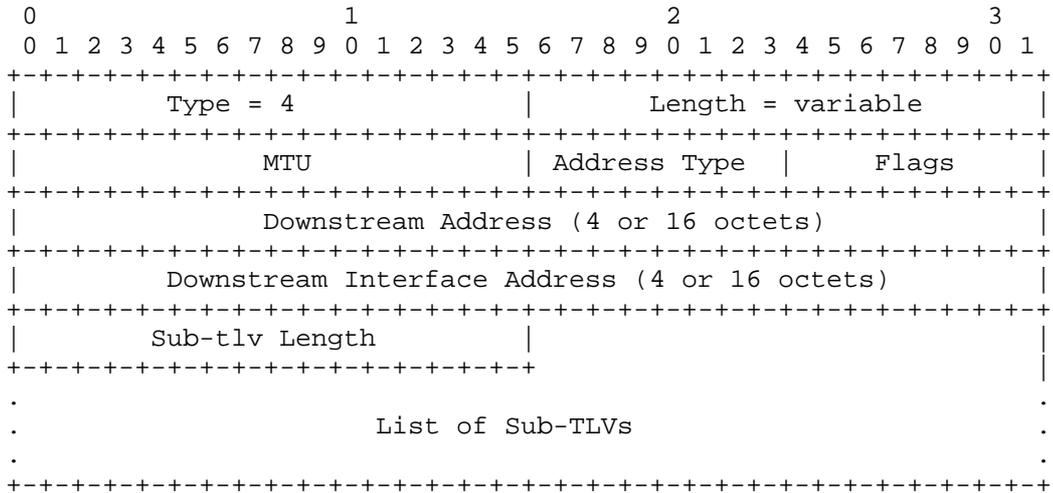
Set ID field is set to the Set Identifier of the incoming BIER-MPLS label. This value is derived as defined in section 2.2 of [I-D.psenak-ospf-bier-extensions]

The BitString field copies the BitString from BIER header of the incoming Echo Request. A Responder BFR SHOULD include this TLV in Echo Reply if the Echo Request is received with I flag set in Downstream Mapping TLV.

An Initiator MUST NOT include this TLV in Echo Request.

3.3.4. Downstream Mapping TLV

This TLV has the following format:



MTU

Set to MTU value of outgoing interface.

Address Type

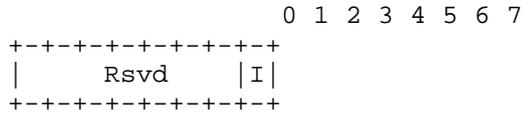
The Address Type indicates the address type and length of IP address for downstream interface. The Address type is set to one of the below:

Type	Addr. Type	DA Length	DIA Length
1	IPv4 Numbered	4	4
2	IPv4 Unnumbered	4	4
3	IPv6 Numbered	16	16
4	IPv6 Unnumbered	16	4

DA Length - Downstream Address field Length
DIA Length - Downstream Interface Address field Length

Flags

The Flags field has the following format:



When I flag is set, the Responding BFR SHOULD include the Incoming SI-BitString TLV in echo reply message.

Downstream Address and Downstream Interface Address

If the Address Type is 1, the Downstream Address MUST be set to IPV4 BFR-Prefix of downstream BFR and Downstream Interface Address is set to downstream interface address.

If the Address Type is 2, the Downstream Address MUST be set to IPV4 BFR-Prefix of downstream BFR and Downstream Interface Address is set to the index assigned by upstream BFR to the interface.

If the Address Type is 3, the Downstream Address MUST be set to IPV6 BFR-Prefix of downstream BFR and Downstream Interface Address is set to downstream interface address.

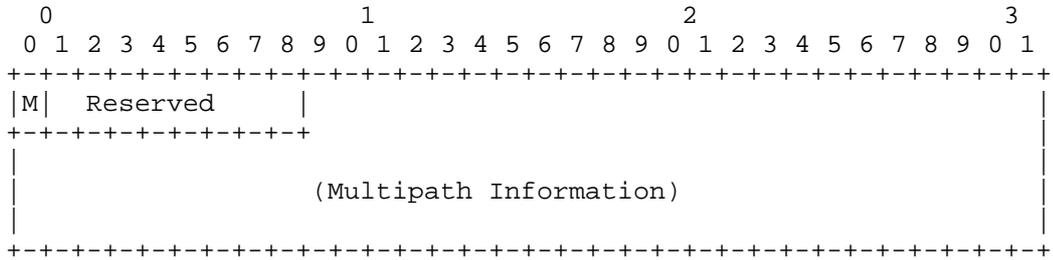
If the Address Type is 4, the Downstream Address MUST be set to IPV6 BFR-Prefix of downstream BFR and Downstream Interface Address is set to index assigned by upstream BFR to the interface.

3.3.4.1. Downstream Mapping Sub-TLVs

This section defines the optional Sub-TLVs that can be included in Downstream Mapping TLV.

Sub-TLV Type	Value
1	Multipath Entropy Data
2	Egress BitString

3.3.4.1.1. Multipath Entropy Data



M Flag

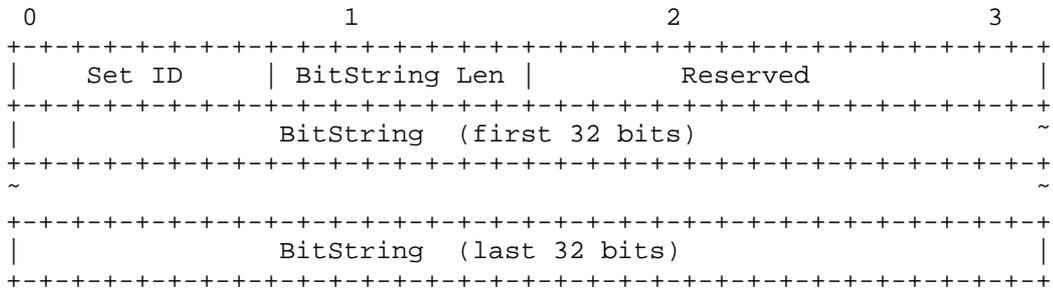
This flag is set to 0 if all packets will be forwarded out through interface defined in Downstream Mapping TLV. When set to 1, Multipath Information will be defined with Bit masked Entropy data.

Multipath Information

Entropy Data encoded as defined in section x3.

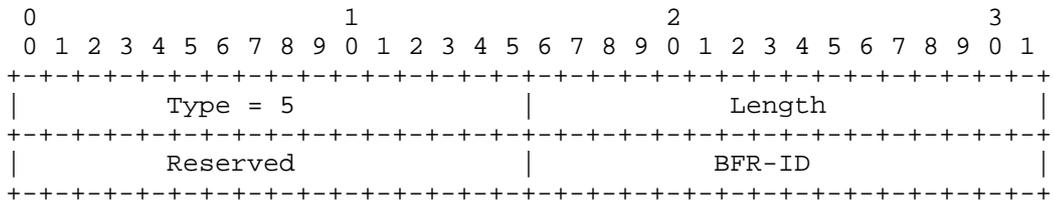
3.3.4.1.2. Egress BitString

This Sub-TLV MAY be included by Responder BFR with the rewritten BitString in outgoing interface as defined in section 6.1 of [I-D.wijnands-bier-architecture]



3.3.5. Responder BFER TLV

The Responder BFER TLV will be included by the BFER replying to the request. This is used to identify the originator of BIER Echo Reply. This TLV have the following format:

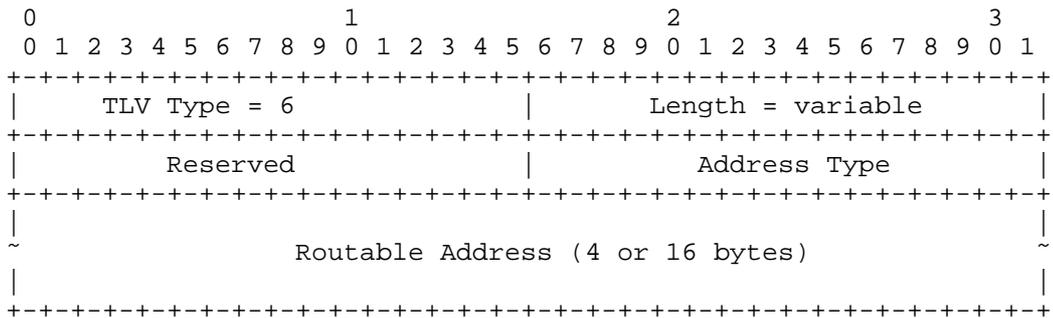


BFR-ID

The BFR-ID field carries the BFR-ID of replying BFER. This TLV MAY be included by Responding BFER in BIER Echo Reply packet.

3.3.6. Responder BFR TLV

The Responder BFR TLV will be included by the transit BFR replying to the request. This is used to identify the replying BFR without BFR-ID. This TLV have the following format:



Length

The Length field varies depending on the Address Type.

Address Type

Set to 1 for IPv4 or 2 for IPv6.

Ratable Address

Carries any locally routable address of replying BFR. This TLV MAY be included by Responding BFR in BIER Echo Reply packet.

- o Presence of Router Alert label in the label stack.

4.2. Per BFER ECMP Discovery

As defined in [I-D.wijnands-bier-architecture], BIER follows unicast forwarding path and allows load balancing over ECMP paths between BFIR and BFER. BIER OAM MUST support ECMP path discovery between a BFIR and a given BFER and MUST support path validation and failure detection of any particular ECMP path between BFIR and BFER.

[I-D.wijnands-mpls-bier-encapsulation] proposes the BIER header with Entropy field that can be leveraged to exercise all ECMP paths. Initiator/BFIR will use traceroute message to query each hop about the Entropy information for each downstream paths. To avoid complexity, it is suggested that the ECMP query is performed per BFER by carrying required information in BIER OAM message.

Initiator MUST include Multipath Entropy Data Sub-TLV in Downstream Mapping TLV. It MUST also include the BFER in BitString TLV to which the Multipath query is performed.

Any transit BFR will reply back with Bit-masked Entropy for each downstream path as defined in [RFC4379]

4.3. Sending BIER Echo Request

Initiator MUST set the Message Type as TBD1 and Return Code as 0. Initiator infer the Set Identifier value from the respective BitString that will be appended in BIER header and include in "SI" field.

The Proto field in OAM packet MUST be set to 0, if there is no data packet following immediately after OAM packet. In all other cases, the Proto field MUST be set to value as defined in [I-D.wijnands-mpls-bier-encapsulation], same as of the data packet that follows after OAM packet.

Initiator MUST include Original SI-BitString TLV. Initiator MUST NOT include more than one Original SI-BitString TLV. In Ping mode, Initiator MAY include Target SI-BitString TLV listing all the BFER from which the Initiator expects a response. In traceroute mode, Initiator SHOULD include Target SI-Bitstring TLV. Initiator on receiving a reply with Return code as "Replying router is one of the BFER in BIER header Bitstring" , SHOULD unset the respective BFR-id from Target SI-Bitstring on any subsequent request.

Initiator MAY also include Downstream Mapping TLV (section 3.3.4). In presence of Multipath Entropy Data Sub-TLV, the Target SI-BitString TLV MUST carry only one BFER id.

Initiator then encapsulates with BIER header and set the Proto as TBD1 and further encapsulates with BIER-MPLS label. In ping mode, the BIER-MPLS Label TTL MUST be set to 255. In traceroute mode, the BIER-MPLS Label TTL is set successively starting from 1 and MUST stop sending the Echo Request if it receives a reply with Return code as "Replying router is the only BFER in BIER header Bitstring" from all BFER listed in BitString TLV.

4.4. Receiving BIER Echo Request

Sending a BIER OAM Echo Request to control plane for payload processing is triggered as mentioned in section 4.1.

Any BFR on receiving Echo Request MUST send Echo Reply with Return Code set to 1, if the packet fails sanity check. If the packet sanity check is fine, it initiates a variable as "Best-return-code" and further processes it as below:

1. Set the Best-return-code to "SI Mismatch", if the received BIER-MPLS Label is not assigned to the Set ID value in Original SI-BitString TLV. Go to section 4.5.
2. Set the Best-return-code to "One or more of the TLVs was not understood", if any of the TLVs in echo request message is not understood. Go to section 4.5.
3. Set the Best-return-code to "Invalid Multipath Info Request", if the Echo Request is received with more than 1 BFER id in Target SI-BitString TLV AND Multipath Entropy Data Sub-TLV. Go to section 4.5.
4. Set the Best-return-code to "Replying router is the only BFER in BIER header Bitstring", and go to section 4.5 if the responder is BFER and there is no more bits in BIER header Bitstring left for forwarding.
5. Set the Best-return-code to "Replying router is one of the BFER in BIER header Bitstring", and include Downstream Mapping TLV, if the responder is BFER and there is more bits in BitString left for forwarding. In addition, include the Multipath information as defined in Section 4.2 if the received Echo Request carries Multipath Entropy Data Sub-TLV. Go to section 4.5.

6. Set the Best-return-code to "No matching entry in forwarding table", if the forwarding lookup defined in section 6.5 of [I-D.wijnands-bier-architecture] does not match any entry for the received BitString in BIER header.
7. Set the Best-return-code to "Packet-Forward-OK", and include Downstream Mapping TLV. Go to section 4.5

4.5. Sending Echo Reply

Responder MUST include DDMAP in Echo Reply if the incoming Echo Request carries DDMAP. Responder MUST set the Message Type as TBD2 and Return Code as Best-return-code. The SI field MUST be set to 0 and Proto field MUST be set to 0.

Responder appends BIER header listing the BitString with BFIR ID (from received Echo Request), set the Proto to PROTO-TBD1 and set the BFIR as zero.

4.6. Receiving Echo Reply

Initiator on receiving echo reply will use the Sender's Handle to match with echo request sent. If no match is found, Initiator MUST ignore the Echo Reply.

If receiving echo reply have Downstream Mapping, Initiator SHOULD copy the same to subsequent Echo Request(s).

5. IANA Considerations

This document request the IANA the creation and management of below registries:

5.1. Message Types, Reply Modes, Return Codes

This document request to assign the Message Types and Reply mode mentioned in section 3.1, , Return code mentioned in Section 3.2.

5.2. TLVs

The TLVs and Sub-TLVs requested by this document for IANA consideration are the following:

Type	Sub-Type	Value Field
-----	-----	-----
1		Original SI-BitString
2		Target SI-BitString
3		Incoming SI-BitString
4		Downstream Mapping
4	1	Multipath Entropy Data
4	2	Egress BitString
5		Responder BFER
6		Responder BFR
7		Upstream Interface

6. Security Considerations

The security consideration for BIER Ping is similar to ICMP or LSP Ping. AS like ICMP or LSP ping, BFR may be exposed to Denial-of-service attacks and it is RECOMMENDED to regulate the BIER Ping packet flow to control plane. A rate limiter SHOULD be applied to avoid any attack.

As like ICMP or LSP Ping, a traceroute can be used to obtain network information. It is RECOMMENDED that the implementation check the integrity of BFIR of the Echo messages against any local secured list before processing the message further

7. Acknowledgement

TBD

8. Contributing Authors

TBD

9. References

9.1. Normative References

[I-D.psenak-ospf-bier-extensions]

Psenak, P., Kumar, N., Wijnands, I., Dolganow, A., Przygienda, T., and J. Zhang, "OSPF Extensions For BIER", draft-psenak-ospf-bier-extensions-02 (work in progress), February 2015.

[I-D.wijnands-bier-architecture]

Wijnands, I., Rosen, E., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast using Bit Index Explicit Replication", draft-wijnands-bier-architecture-04 (work in progress), February 2015.

- [I-D.wijnands-mpls-bier-encapsulation]
Wijnands, I., Rosen, E., Dolganow, A., Tantsura, J., and S. Aldrin, "Encapsulation for Bit Index Explicit Replication in MPLS Networks", draft-wijnands-mpls-bier-encapsulation-02 (work in progress), December 2014.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC4379] Kompella, K. and G. Swallow, "Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures", RFC 4379, February 2006.
- [RFC5905] Mills, D., Martin, J., Burbank, J., and W. Kasch, "Network Time Protocol Version 4: Protocol and Algorithms Specification", RFC 5905, June 2010.

9.2. Informative References

- [RFC0792] Postel, J., "Internet Control Message Protocol", STD 5, RFC 792, September 1981.
- [RFC6291] Andersson, L., van Helvoort, H., Bonica, R., Romascanu, D., and S. Mansfield, "Guidelines for the Use of the "OAM" Acronym in the IETF", BCP 161, RFC 6291, June 2011.
- [RFC6424] Bahadur, N., Kompella, K., and G. Swallow, "Mechanism for Performing Label Switched Path Ping (LSP Ping) over MPLS Tunnels", RFC 6424, November 2011.
- [RFC6425] Saxena, S., Swallow, G., Ali, Z., Farrel, A., Yasukawa, S., and T. Nadeau, "Detecting Data-Plane Failures in Point-to-Multipoint MPLS - Extensions to LSP Ping", RFC 6425, November 2011.

Authors' Addresses

Nagendra Kumar
Cisco Systems, Inc.
7200 Kit Creek Road
Research Triangle Park, NC 27709
US

Email: naikumar@cisco.com

Carlos Pignataro
Cisco Systems, Inc.
7200 Kit Creek Road
Research Triangle Park, NC 27709-4987
US

Email: cpignata@cisco.com

Nobo Akiya
Cisco Systems, Inc.
2000 Innovation Drive
Kanata, ON K2K 3E8
Canada

Email: nobo@cisco.com

Lianshu Zheng
Huawei Technologies
China

Email: vero.zheng@huawei.com

Mach Chen
Huawei Technologies

Email: mach.chen@huawei.com

Greg Mirsky
Ericsson

Email: gregory.mirsky@ericsson.com